

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 09/936,823A
Source: 1FW16
Date Processed by STIC: 10/12/05

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 10/12/2005

PATENT APPLICATION: US/09/936,823A

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

3 <110> APPLICANT: Merja PENTTILA et al.
 5 <120> TITLE OF INVENTION: PROCESS FOR PARTITIONING OF PROTEINS
 7 <130> FILE REFERENCE: 0933-0170P
 9 <140> CURRENT APPLICATION NUMBER: US 09/936,823A
 10 <141> CURRENT FILING DATE: 2001-10-24
 12 <150> PRIOR APPLICATION NUMBER: PCT/FI00/00249
 13 <151> PRIOR FILING DATE: 2000-03-24
 15 <150> PRIOR APPLICATION NUMBER: FI 19991782
 16 <151> PRIOR FILING DATE: 1999-08-20
 18 <150> PRIOR APPLICATION NUMBER: FI 19990667
 19 <151> PRIOR FILING DATE: 1999-03-25
 21 <160> NUMBER OF SEQ ID NOS: 46
 23 <170> SOFTWARE: PatentIn Ver. 2.2
 25 <210> SEQ ID NO: 1
 26 <211> LENGTH: 428
 27 <212> TYPE: DNA
 28 <213> ORGANISM: Trichoderma reesei
 30 <220> FEATURE:
 31 <221> NAME/KEY: intron
 32 <222> LOCATION: (167)..(236)
 34 <220> FEATURE:
 35 <221> NAME/KEY: intron
 36 <222> LOCATION: (323)..(386)
 38 <220> FEATURE:
 39 <221> NAME/KEY: misc_feature
 40 <223> OTHER INFORMATION: Coding sequence of hfb1
 42 <400> SEQUENCE: 1
 43 atgaagtcttcgcccattcgc cgctctttt gccggcgctg ccgttgccca gcctctcgag 60
 44 gaccgcagca acggcaacgg caatgttgc cctccggcc tcttcagcaa cccccagtgc 120
 45 tgtgccaccc aagtccctgg cctcatggc ctgtactgca aagtccgtaa gttgagccat 180
 46 aacataagaa tcctcttgac ggaaatatgc ctctcactc ctttaccctt gaacagccctc 240
 47 ccagaacgtt tacgacggca cggactccg caacgtctgc gccaaaaccg ggcggccaggcc 300
 48 tctctgtgc gtggcccccgg ttgttaagggtt atgccccaggc tcaagctcca gtctttggca 360
 49 aacccttctt gacacccaga ctgcaggccg gccaggctct tctgtgccag accggccgtcg 420
 50 gtgttga
 428
 53 <210> SEQ ID NO: 2
 54 <211> LENGTH: 78
 55 <212> TYPE: DNA
 56 <213> ORGANISM: Artificial Sequence
 58 <220> FEATURE:
 59 <223> OTHER INFORMATION: Description of Artificial Sequence: PCR 5' primer
 61 <400> SEQUENCE: 2
 62 tcgggcacta cgtgccagta tagcaacgac tactactcgc aatgccttgt tccgcgtggc 60

P.6

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/936,823A

DATE: 10/12/2005

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT
 Output Set: N:\CRF4\10122005\I936823A.raw

```

63      tctagttctg gaaccgca                                78
66 <210> SEQ ID NO: 3
67 <211> LENGTH: 30
68 <212> TYPE: DNA
69 <213> ORGANISM: Artificial Sequence
71 <220> FEATURE:
72 <223> OTHER INFORMATION: Description of Artificial Sequence: PCR 3' primer
74 <400> SEQUENCE: 3
75      tcgtacggat cctcaaggcac cgacggcggt                               30
78 <210> SEQ ID NO: 4
79 <211> LENGTH: 63
80 <212> TYPE: DNA
81 <213> ORGANISM: Artificial Sequence
83 <220> FEATURE:
84 <223> OTHER INFORMATION: Description of Artificial Sequence: PCR 5' primer
86 <400> SEQUENCE: 4
87      actacacgga ggagctcgac gacttcgagc agcccgagct gcacgcagag caacggcaac 60
88      ggc                                         63
91 <210> SEQ ID NO: 5
92 <211> LENGTH: 2211
93 <212> TYPE: DNA
94 <213> ORGANISM: Trichoderma reesei
96 <220> FEATURE:
97 <221> NAME/KEY: promoter
98 <222> LOCATION: (1)..(2211)
99 <223> OTHER INFORMATION: cbh1 promoter sequence
101 <400> SEQUENCE: 5
102     gaatttcac ggtgaatgtt ggcctttgtt agggtaggaa ttgtcactca agcacccccca 60
103     acctccatta cgcctcccccc atagagttcc caatcagtga gtcatggcac tttctcaaa 120
104     tagattgggg agaagggttgc ttccggccag agctgaagggt cgccacaaccg catgatatacg 180
105     ggtcgccaac ggcaaaaaaaag cacgtggctc accgaaaagc aagatgtttt cgatctaaca 240
106     tccaggaacc tgatatacatc catcatcactc cacgaccact ttgatctgtt ggtaaactcg 300
107     tattcgccctt aaaccgaagt gctgtttaaa tctacacgtt ggccttttc ggtataactgc 360
108     gtgtgtttc tcttaggttca ttcttcctt cctcttaggtt tgaattgtttt gtgtggag 420
109     tccgagctgtt aactacctctt gaatctctgg agaatgggtt actaacgact accgtgcacc 480
110     tgcatacatgtt atataatagt gatccttgaga aggggggtttt ggagcaatgtt gggactttga 540
111     tggtcatcaa accaagaacg aagacgcctc ttttgcaaaat ttttgttgc gctacgggtga 600
112     agaactggat acttggatgtt tcttcgtgtt atttttgtgg caacaagagg ccagagacaa 660
113     tctattcaaa caccaagctt gctcttttga gctacaagaa cctgtgggtt atatatctgtt 720
114     agttgttgaag tcggtaatcc cgctgtatag taatacgtt cgcatctaaa tactccgaag 780
115     ctgctgcgaa cccggagaat cgagatgtgc tggaaagctt ctgcgcgcg gctaaatttt 840
116     catgaaaggc tatgagaaat tctggagacg gcttgggtt gatggcgat ccattcttcg 900
117     acaagcaaaag cgttccgtcg cagtagcagg cactcattcc cggaaaaactt cggagatccc 960
118     taatgttgcgaa tggaaaccggaa ataataat aggcaatataca ttgagttgcc tcgcacgggtt 1020
119     caatgcagggtt gtactgagct tggacataac tggatgttccgtac cccacccctt ctcaacccctt 1080
120     ggcgtttccctt tgattcagcg taccgttaca agtctgttactt actattaaacc cagactgacc 1140
121     ggacgtgtttt tggatgttcat tggatgttactt aatgttgcattt cgtatgtttaa tttgcctgtt 1200
122     tgaccgttgc gggctgttgc aagccgttgc aatgttgcattt gtaggttgcattt tttgcctgtt 1260
123     aggcatgttgc tgaatctgttgc tggggcaggaa caccgcctcgaa aggttgcacgg caaggaaac 1320

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/936,823A

DATE: 10/12/2005

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

124 caccgatagc agtgtctagt agcaacctgt aaagccgcaa tgcagcatca ctggaaaata 1380
 125 caaaccatg gctaaaagta cataagttaa tgcctaaaga agtcatatac cagcggctaa 1440
 126 taattgtaca atcaagtggc taaacgtacc gtaatttgc aacgcgttgc ggggttcag 1500
 127 aagcaacggc aaagcccact tcccacgtt gtttctcac tcagtc当地 1560
 128 gatccccaa ttgggtcgct tgtttgc ggtgaagtga aagaagacag aggttaagaat 1620
 129 gtctgactcg gagcgttttgc catacaacca agggcagtga tggaaagacag taaaatgtt 1680
 130 acattcaagg agtatttagc cagggatgct tgagtgtatc gtgttaaggag gtttgc当地 1740
 131 cgatacgacg aatactgtat agtcaattctt gatgaagtgg tccatattga aatgttaagtc 1800
 132 ggcactgaac aggcaaaaaga ttgaggtaa actgcctaag atctcgggccc ctcgggcttc 1860
 133 ggctttgggt gtacatgtt gtgtccggg caaatgcaaa gtgtggtagg atcgacacac 1920
 134 tgctgcctt accaaggcagc tgagggtatg tgataggcaa atgttcaggg gccactgcat 1980
 135 ggttcaat agaaagagaa gcttagccaa gaacaatagc cgataaaagat agcctcatta 2040
 136 aacgaaatga gctagtaggc aaagtccagc aatgtgtata tataaaagggtt cgaggtccgt 2100
 137 gcctccctca tgctctcccc atctactcat caactcagat cctccaggag acttgtacac 2160
 138 catctttga ggcacagaaaa cccaatagtc aaccgcggac tgcgc当地cat 2211
 141 <210> SEQ_ID NO: 6
 142 <211> LENGTH: 1588
 143 <212> TYPE: DNA
 144 <213> ORGANISM: Trichoderma reesei
 146 <220> FEATURE:
 147 <221> NAME/KEY: misc_feature
 148 <223> OTHER INFORMATION: T. reesei egl1 cDNA
 150 <400> SEQUENCE: 6
 151 cccccctatc tttagtccttc ttgttgtccc aaaatggcgc cctcagttac actgc当地ttg 60
 152 accacggcca tcctggccat tgccggctc gtgc当地ggccc agcaacccggg taccagcacc 120
 153 cccgaggtcc atcccaagtt gacaacctac aagtgtacaa agtccggggg gtgc当地ggcc 180
 154 caggacaccc cgttggctt tgactggAAC taccgtcttga tgc当地cgc当地cc 240
 155 tcgtgcaccc tcaacggcgg cgtcaacacc acgctcttgc当地 ctgac当地ggc gacctgtggc 300
 156 aagaactgct tcatcgaggg cgtc当地actac gccgc当地tgg gctc当地acgc当地 ctc当地ggc当地 360
 157 agcctc当地cca tgaaccaggta catgccc当地c agctcttggc当地 gctacagcagc cgtctcttcc 420
 158 cggctgtatc tcctggactc tgacggtgag tacgtgtatc tgaagctcaa cggccaggag 480
 159 ctgagcttcg acgtcgaccc ctctgtcttgc当地 ccgtgtggag agaacggc当地 gctctaccc 540
 160 tctcagatgg acgagaacgg gggc当地caac cagtataaca cggcc当地gtgc当地 caactacggg 600
 161 agcggctact gcatgctca gtgc当地ggc当地 cagacatggc当地 ggaacggc当地 cctcaacact 660
 162 agccaccagg gcttctgtc当地 caacgagatg gatatccttgc当地 agggcaactc gagggc当地aat 720
 163 gccttgaccc ctcaacttgc当地 cacggccaccc gcctglocalact ctgccgggtt cggcttcaac 780
 164 ccctatggca gccc当地tacaa aagctactac gccccggag ataccgttgc当地 caccttcaac 840
 165 accttccatca tcatcaccatca gttcaacaccg gacaacccggc cgc当地cttggg caaccttgc当地 900
 166 agcatcaccc gcaagtgatcca gcaaaaacggc gtgc当地acatcc ccagc当地ccca gccc当地ggc当地 960
 167 gacaccatct cgtc当地tgc当地 gtc当地ccctca gcctacggc当地 gcctglocalcc catgggcaac 1020
 168 gccctgagca gccc当地atggc当地 gctc当地gttgc当地 agcatttgc当地 acgacaacag ccagtgatcg 1080
 169 aactggctcg acagcggca gccc当地ggcc当地 tgc当地agc当地 cc当地aggccaa cccatccaa 1140
 170 atccctggcca acaacccca cacgc当地acgtc gtcttcttca acatccgctg gggagacatt 1200
 171 gggtctacta cgaactcgac tgc当地ccccc当地 cccccc当地tgc当地 cgtccagc当地 gacgtt当地tc当地 1260
 172 actacacggc ggagctcgac gacttc当地gagc agccc当地gagct gc当地acgc当地ac tcactggggg 1320
 173 cagtgc当地ggc当地 gcatgggta cagc当地gggtc当地 aagacgttgc当地 cgtccggc当地 tacgtglocalcc 1380
 174 tatagcaacg actactactc gcaatgc当地tt tagagc当地ttg acttgc当地ctc ggtctgttca 1440
 175 gacggggca cgatagaatg cgggc当地acgc当地 gggagc当地gtg agacatttggg cttatataat 1500
 176 aagacatgct atgttgc当地tacat tagca aatgacaac aaatgaaaaa gaacttatca 1560

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/936,823A

DATE: 10/12/2005

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT
 Output Set: N:\CRF4\10122005\I936823A.raw

```

177      agcaaaaaaa aaaaaaaaaa aaaaaaaaaa                               1588
180 <210> SEQ ID NO: 7
181 <211> LENGTH: 745
182 <212> TYPE: DNA
183 <213> ORGANISM: Trichoderma reesei
185 <220> FEATURE:
186 <221> NAME/KEY: terminator
187 <222> LOCATION: (1)..(745)
188 <223> OTHER INFORMATION: T. reesei cbh1 terminator
190 <400> SEQUENCE: 7
191      ggacctaccc agtctacta cggccagtgc ggcggattg gctacagcgg ccccacggtc 60
192      tgcgccagcg gcacaacttg ccaggtcctg aacccttaact acttcagtg cctgtaaagc 120
193      tccgtgcgaa agcctgacgc accggtagat tcttggtgag cccgtatcat gacggcgccg 180
194      ggagctacat ggccccgggt gatttatttt ttttgtatct acttctgacc ctttcaaat 240
195      atacggtcaa ctcatcttc actggagatg cggcctgctt ggtattgcga tggtgtcagc 300
196      ttggcaaatt gtggctttcg aaaacacaaa acgattccctt agtagccatg catttaaga 360
197      taacggaata gaagaaaagag gaaattaaaa aaaaaaaaaa aacaaacatc ccgttcataa 420
198      cccgtagaat cggcgctctt cgtgtatccc agtaccacgt caaaggattt catgatcgtt 480
199      caatgttcat attgttccgc cagttggctt ccaccccat ctccgcgaat ctccctttct 540
200      cgaacgcgggt agtggctgctt gccaattggg aatgaccata gggagacaaa cagcataata 600
201      gcaacagtgg aaatttagtgg cgcaataatt gagaacacag tgagaccata gctggcgcc 660
202      tggaaagcac tggggagac caacttgtcc gttgcgaggc caacttgcatt tgctgtcaag 720
203      acgatgacaa cgtagccgag gaccc                                         745
206 <210> SEQ ID NO: 8
207 <211> LENGTH: 10
208 <212> TYPE: DNA
209 <213> ORGANISM: Artificial Sequence
211 <220> FEATURE:
212 <223> OTHER INFORMATION: Description of Artificial Sequence: annealed primer
214 <400> SEQUENCE: 8
215      taaccgcggt                                         10
218 <210> SEQ ID NO: 9
219 <211> LENGTH: 16
220 <212> TYPE: DNA
221 <213> ORGANISM: Artificial Sequence
223 <220> FEATURE:
224 <223> OTHER INFORMATION: Description of Artificial Sequence: annealed primer
226 <400> SEQUENCE: 9
227      ctagaccgcg gttaat                                         16
230 <210> SEQ ID NO: 10
231 <211> LENGTH: 1232
232 <212> TYPE: DNA
233 <213> ORGANISM: Trichoderma reesei
235 <220> FEATURE:
236 <221> NAME/KEY: promoter
237 <222> LOCATION: (1)..(1232)
238 <223> OTHER INFORMATION: T. reesei gpdl promotor
240 <400> SEQUENCE: 10
241      gtcgacacga tatacaggcg cggctgatga taatgtatcgat cgagcatgac ttgatgtgt 60

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/936,823A

DATE: 10/12/2005

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

242 atgtgacaat attgactgcg aggaaccatc aggtgttat gnatgaaatc attctgtac 120
 243 caccaggta catgcatacat aaggattctc ctcagctcac caacaacgaa cgatggccat 180
 244 gtttagtgaag gcaccgtat ggcaagatag aaccactatt gcatctgcgc ttcccacgca 240
 245 cagtacgtca agtaaacgtca aagcccccct cccgttaacct cggccgttgt tgctcccccc 300
 246 gattgcctca atcacatagt acctacctat gcattatggg cggcctcaac ccacccccc 360
 247 agattgagag ctaccttaca tcaatatggc cagcacctt tcggcgatac atactcgcca 420
 248 ccccagccgg cgcgattgtg tgtacttagt aggctcgatc tataccagca ggagagggtgc 480
 249 tgcttggcaa tcgtgtcttagt ctgttagtt gtacttgat ggtacttgta aggtggtcat 540
 250 gcagttgtca agttacccatggg ggagggattc aacgagccct gtttccatg tccatctgga 600
 251 taggatggcg gctggcgcccccc cccgaagctgg gaactcgcca acagtcatat gtaatagctc 660
 252 aagttgtatga taccgttttgc ccagattaga tgcgagaagc agcatgaatg tcgctcatcc 720
 253 gatgccgcat caccgttgc tcagaaacga ccaagctaaag caactaaggt accttaccgt 780
 254 ccactatctc aggttaaccatggg gtactaccatggg ctaccctacc tcggcgccct acctgctta 840
 255 gtgttaatct ttccacctcc ctcctcaatc ttcttttccc tcctctctc tttttttttt 900
 256 ctccctctc ttcttctcca taaccattcc taacaacatc gacattctct cctaattcacc 960
 257 agcctcgcaa atcctcagtt tgtatgtacg tacgtactac aatcatcacc acgatcgcc 1020
 258 gcccgcacat gcccgcacat ttgcctgccc ctcctctca ctcgtgcct tgacgagcta 1080
 259 gccccgcacat gactctcctg cgtcaccat tttttccctt atttaccctt ctcctctc 1140
 260 tccctctcgt ttcttcctaa caaacaacca ccacccaaat ctcttggaa gtcacgact 1200
 261 cacgcagtc aattcgcaga tacaaatcta ga 1232

264 <210> SEQ ID NO: 11
 265 <211> LENGTH: 1129
 266 <212> TYPE: DNA
 267 <213> ORGANISM: Trichoderma reesei
 269 <220> FEATURE:
 270 <221> NAME/KEY: terminator
 271 <222> LOCATION: (1)..(1129)
 272 <223> OTHER INFORMATION: T. reesei gpd1 terminator
 274 <400> SEQUENCE: 11

275 ggatcccgag cattgtctat gaatgaaac aaaaatagta aataaatagt aattctggcc 60
 276 atgacgaata gagccaatct gtcacttgc actatcttg gactgtatcg tatgtcgaaac 120
 277 ccttgcactgc ccattcaac aattgtaaag gaatatagtt acaagttatg ttcacatgttt 180
 278 gcgtgcgagc ccgtttgtac gttatgttgc gaaagcggtt ccacatcatg ctcacatgtca 240
 279 cttggcttac gatcatgttt gcgatctcg gtaagaatac acagagtaac gattatctcc 300
 280 atcgcttcta tgatttagtta ctcagacaac acatgggaaa caagataacc atcgcatgca 360
 281 aggtcgattt caatcatgtat ctggactggg gtattccatc taagccatag taccctcgag 420
 282 agaaggaaatg gttaggacctc tcaggcgcc accatctgtg ctgcaaattcc aagaaacccc 480
 283 cccaaagcac ctacctatct accttagatgta actgcacgag aaaagaaaag gagcagaaga 540
 284 agaatgatct caagaggccg tgaacgcaga aacacactcc tcccaacttt tcaagtttg 600
 285 aacaaaaaaa gaaagatgag gactagaaga tggagtattt ctttctttaga gagctctcg 660
 286 tgaggtgacc tgcagggtt taccgaaac cgtcggttgt tctatccat taatcaagtc 720
 287 cccgcgcctcg cctcttctct cctgtctttt catagaatcc cgtctccttg ttgcttgatc 780
 288 gaagcgccgtt tatecgacgccc accaaagatc ttgtcttgat gacttataa tccttgggt 840
 289 atcaaacacgc ccccgagtga tcagatccgt aaaagaagaa gaagagtacg attaaccag 900
 290 accgaggaac aataaaagcgta gtaaaaataaca tcaaaaataag agtctcggtt aaaattactt 960
 291 gttcctcaat caatcccaac cccctaaaaa gcccctcccccc ccatggata tcccgccagt 1020
 292 aggagagaga tatttccact accgctcacc accaagttagt gcttggcgag agaagaggat 1080
 293 gaatcagaag tgacaacaac gggttgagca catggatcatggat cggcgcc 1129

296 <210> SEQ ID NO: 12

6

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 10/12/2005
PATENT APPLICATION: US/09/936,823A TIME: 11:23:20

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT
Output Set: N:\CRF4\10122005\I936823A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:12; N Pos. 2570,2778,2869

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:1; Line(s) 1

Seq#:42; Line(s) 1095

VERIFICATION SUMMARY DATE: 10/12/2005
PATENT APPLICATION: US/09/936,823A TIME: 11:23:20

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT
Output Set: N:\CRF4\10122005\I936823A.raw

L:368 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:2520
M:341 Repeated in SeqNo=12